Time series integrative analysis of RNA-Seq and miRNA expression data reveals key biologic wound healing pathways in keloid-prone individuals

Alexandros Onoufriadis[1,*], Chao-Kai Hsu[2,3,*], Chrysanthi Ainali[1], Chuin Ying Ung[1], Ellie Rashidghamat[1], Hsing-San Yang[2], Hsin-Yu Huang[2,4], Umar Niazi[5], Christos Tziotzios[1], Jui-Chu Yang[6], Rosamond Nuamah[7], Ming-Jer Tang[3,8], Alka Saxena[7], Emanuele de Rinaldis[5], John A. McGrath[1]

[1]St John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London, London, United Kingdom

[2]Department of Dermatology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[3]International Research Center of Wound Repair and Regeneration (iWRR) National Cheng Kung University, Tainan, Taiwan

[4]School of Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[5]Guy's and St Thomas' NHS Foundation Trust and King's College London NIHR Biomedical Research Centre Translational Bioinformatics Platform, Guy's Hospital, London, United Kingdom

[6]Human Biobank, Research Centre of Clinical Medicine, National Cheng Kung University Hospital, Tainan, Taiwan

[7]Guy's and St Thomas' NHS Foundation Trust and King's College London NIHR Biomedical Research Centre Genomics Research Platform, Guy's Hospital, London, United Kingdom

[8]Department of Physiology, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[*]These authors contributed equally to this work

Corresponding author: John A. McGrath, Dermatology Research Labs, Floor 9 Tower Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, United Kingdom. Tel: 44-2071886409; Fax: 44-2071888050; E-mail: john.mcgrath@kcl.ac.uk

Short title: Early keloid transcriptome

Abbreviations used: RNA-Seq, RNA sequencing; miRNA, microRNA; MAPK, mitogen-activated protein kinase; DE, differentially expressed; GSVA, gene set variation analysis; GSEA, gene set enrichment analysis; IL, interleukin

Funding sources: The authors acknowledge financial support from The Rosetrees Trust (M348) and by the grants of the Ministry of Science and Technology, Taiwan (MOST105-2320-B-006-013).

Conflicts of interest: None declared

Word count: 1,056

Figures: 2

Keywords: Keloid, scarring, wound healing, transcriptome

Keloidal scarring is a common and disfiguring skin problem yet its pathobiology is only partially understood and treatments remain sub-optimal (Glass 2017). To date, most

investigative studies have focused on established keloid lesions and the surrounding extracellular matrix (He et al., 2017). In contrast, here we explored transcriptomic alterations at an earlier time-point – during keloid formation. We studied keloid-prone individuals from pedigrees with an autosomal dominant history of keloids, as well as unaffected family members and healthy matched control subjects without any tendency to form keloids (see Supplementary Figure S1). All subjects were Taiwanese.

Following institutional ethics approval and written informed consent, we performed 3mm punch biopsies of non-lesional upper outer buttock skin, followed by a further 4mm punch biopsy of the same site 6 weeks later (see Supplementary Table S1 and Supplementary Figure S2). For the study, biopsying buttock skin was deemed acceptable by both the participants and the ethics' committee (see Supplementary material for further discussion thereof). The 6-week time-point was chosen based on feedback from the keloid-prone individuals as to when they were normally first aware that a keloid scar was developing. We undertook an integrative approach of RNA-Seq and miRNA expression analysis based on the two sets of skin biopsies (baseline and 6 weeks later).

The study involved 8 keloid-prone subjects and 6 healthy matched individuals. Each skin biopsy was immediately immersed in RNAlater (Thermo Fisher Scientific) and total RNA was isolated using the RNeasy Plus Universal kit (Qiagen), retaining miRNAs according to the manufacturer's protocol. RNA samples were subjected to microarray analysis on Affymetrix GeneChip miRNA 4.0 arrays and total RNA-Seq analysis on Illumina pair-end sequencing (see Supplementary Materials). The RNA-Seq raw data files and metadata have been deposited in the Sequence Read Archive (SRA ID: SRP137071) and the miRNA raw data and metadata in Gene Expression Omnibus (GEO ID: GSE113621).

A stepwise bioinformatics strategy was followed to identify differentially expressed miRNAs that may contribute to keloid pathogenesis (see Supplementary Materials and Supplementary Figure S3). This analysis highlighted 37 miRNAs that were differentially expressed in the keloid-prone subjects. Hierarchical clustering revealed two clusters which were upregulated 6 weeks after wounding (see Supplementary Figure S4).

In parallel, differentially expression (DE) analysis was applied to the RNA-Seq data between keloid-prone and healthy subjects, which identified 8 genes at baseline and 47 genes at 6 weeks after wounding that were differentially expressed (adjusted P-value < 0.05; see Supplementary Materials). Comparing healthy controls before and after wounding identified 2,215 differentially expressed genes, whereas the same analysis in the keloid-prone individuals identified 3,161 differentially expressed genes (see Supplementary Figure S5a). Of those genes, there were 513 genes specific to the healthy individuals and 1,449 genes specific to the keloid phenotype (see Supplementary Figure S5b). Hierarchical clustering of the differentially expressed genes specific to the keloid phenotype exhibited two distinct clusters showing changes in expression between baseline and 6 weeks after wounding (see Supplementary Figure S6).

We further assessed pathway enrichment in the RNA-Seq data using the Gene Set Variation Analysis (GSVA) package (see Supplementary Figure S7) (Hänzelmann et al, 2013). For genes specific to the keloid phenotype, there were 101 differentially activated pathways between baseline and 6 weeks after wounding, while 24 pathways were found to be differentially activated for the genes that were specific to the healthy individuals (Figure 1, and see Supplementary Tables S2-S3). Of these, 22 pathways

that were specific to the keloid-prone individuals were present on the KEGG and Reactome pathway databases, which are manually curated and well-annotated.

Of note, NOTCH signaling, MAPK signaling and Toll-like receptor pathways were found to be altered in keloid-prone individuals after wounding with a decrease in pathway activity. These pathways have already been suggested to play a role in keloid disease, and thus our analysis provides further evidence to support their involvement (Bagabir et al., 2011; Syed and Bayat., 2012; Wu et al., 2017). Moreover, DNA repair and p53 signaling pathways were also highlighted (Yamauchi et al., 2018). In addition, the analysis also identified altered regulation of insulin secretion and metabolic pathways (RNA, protein, fructose, mannose and glycerophospholipid metabolism) in keloid pathobiology. Of note, recent work has shown increased glycolytic metabolism in keloid fibroblasts suggesting that dysregulation of metabolic pathways such as glucose metabolism may contribute to keloid formation (Li et al., 2018).

To identify the targetome of the differentially expressed miRNAs for each of the two clusters of the differentially expressed genes in keloid-prone individuals (see Supplementary Figures S4 and S6), we intersected the 37 miRNAs with the 1,449 genes that were specific to the keloid phenotype and that were identified from the analyses described above. As a result, there were 403 over-expressed mRNA-miRNA interactions for 24 DE miRNAs for cluster 1 and 635 down-regulated mRNA-miRNA interactions for 29 DE miRNAs. Figure 2a visualizes the networks derived from both up- and down-regulated putative targets that are specific to the keloid phenotype 6 weeks after wounding.

Next, to investigate the functional dynamic changes of the 1,449 DE genes that were involved in the targetome we conducted gene set enrichment analysis (GSEA) using the R package GAGE (Luo et al., 2009). This analysis identified MAPK signaling

pathway as the only gene set to be significantly dysregulated in the keloid prone subjects 6 weeks after wounding (see Supplementary Tables S4-S5). Notably, other published data have shown that inhibition of MAPK hinders invasive growth of keloid fibroblasts (Wu et al., 2017).

Gene association network analysis was also performed to further classify the differentially expressed genes from the RNA-Seq dataset according to Reactome pathway terms and correlated expression values amongst them (Figure 2b, and see Supplementary Materials). This analysis demonstrated a divergent average expression profile of cytokine signaling genes between keloid-prone and healthy individuals during wound healing. Of note, interleukin (IL)-1α, IL-1β, IL-6, and TNF-α proinflammatory cytokines have been shown to be upregulated in keloid tissue (Ogawa et al., 2016). Differences in organelle biogenesis and metabolism were also highlighted, providing further support that dysregulation of metabolic pathways may contribute to keloid formation.

In summary, our study provides a comprehensive and integrative analysis of the keloid transcriptome and miRNAome and highlights biological pathways that feature during keloid formation. Functional validation will be required to confirm these findings and determine mechanistic and potential therapeutic relevance. Similar studies at earlier time-points after wounding are also likely to add further insight to keloid biogenesis.

Acknowledgments

This work was supported by the grants of the Ministry of Science and Technology, Taiwan (MOST105-2320-B-006-013), The Rosetrees Trust (M481) and by the UK National Institute for Health Research (NIHR) comprehensive Biomedical Research

Centre (BRC) award to Guy's and St. Thomas' NHS Foundation Trust, in partnership with the King's College London and King's College Hospital NHS Foundation Trust. We acknowledge statistical assistance by Dr Lazaros Mavridis and Dr Tanya Shaw for her comments on the manuscript. We would also like to acknowledge all the patients, their relatives, and volunteers who have kindly contributed samples.

References

Bagabir RA, Syed F, Rautemaa R, McGrouther DA, Paus R, Bayat A. Upregulation of Toll-like receptors (TLRs) 6, 7, and 8 in keloid scars. J Invest Dermatol 2011;131:2128-30.

Bran GM, Goessler UR, Hormann K, Riedel F, Sadick H. Keloids: current concepts of pathogenesis (review). Int J Mol Med 2009;24:283-93

Glass DA 2nd. Current Understanding of the Genetic Causes of Keloid Formation. J Investig Dermatol Symp Proc 2017;18:S50-S53.

Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.

He Y, Deng Z, Alghamdi M, Lu L, Fear MW, He L. From genetics to epigenetics: new insights into keloid scarring. Cell Prolif 2017;50.

Li Q, Qin Z, Nie F, Bi H, Zhao R, Pan B, et al. Metabolic reprogramming in keloid fibroblasts: Aerobic glycolysis and a novel therapeutic strategy. Biochem Biophys Res Commun 2018;496:641-647.

Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 2009;10:161.

Ogawa R. Keloid and hypertrophic scars are the result of chronic inflammation in the reticular dermis. Int J Mol Sci 2017;18:606.

Syed F, Bayat A. Notch signaling pathway in keloid disease: enhanced fibroblast activity in a Jagged-1 peptide-dependent manner in lesional vs. extralesional fibroblasts. Wound Repair Regen 2012;20:688-706.

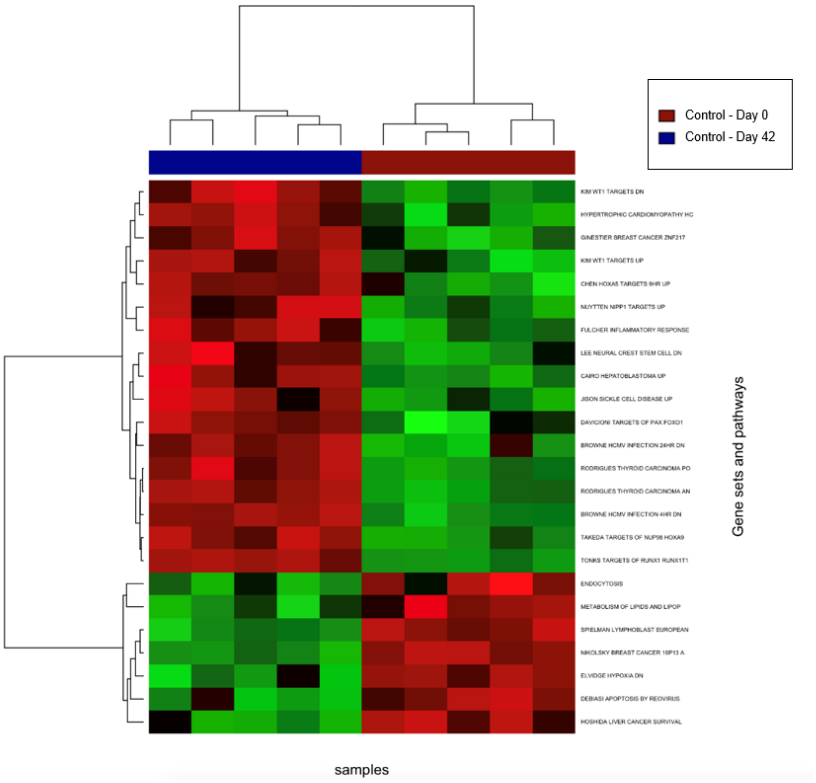
Wu X, Bian D, Dou Y, Gong Z, Tan Q, Xia Y, et al. Asiaticoside hinders the invasive growth of keloid fibroblasts through inhibition of the GDF-9/MAPK/Smad pathway. J Biochem Mol Toxicol 2017;31.

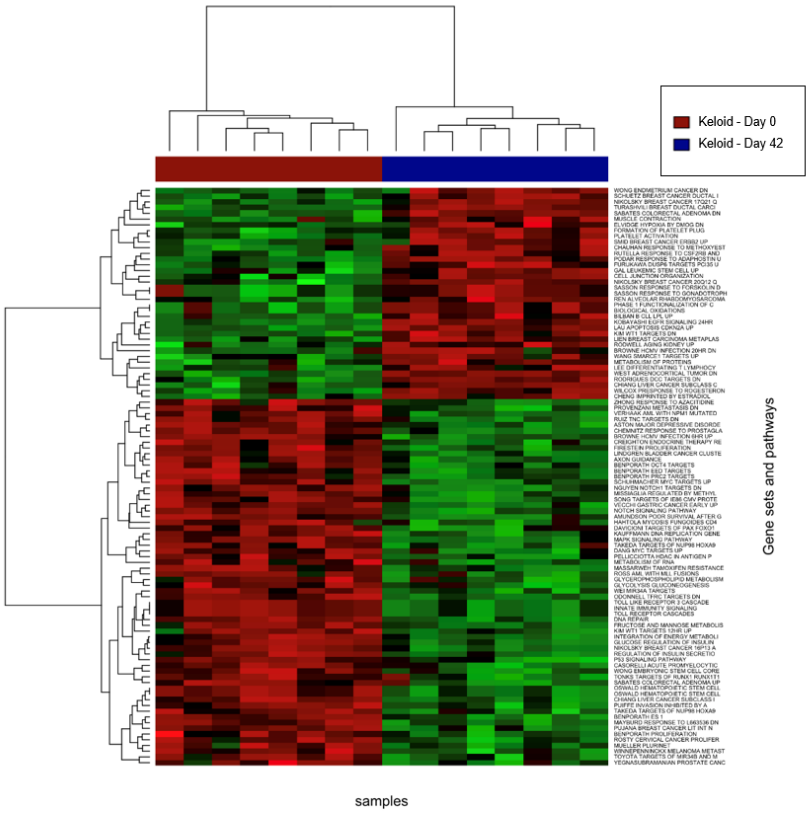
Yamauchi M, Barker TH, Gibbons DL, Kurie JM. The fibrotic tumor stroma. J Clin Invest 2018;128:16-25.

Figure legends

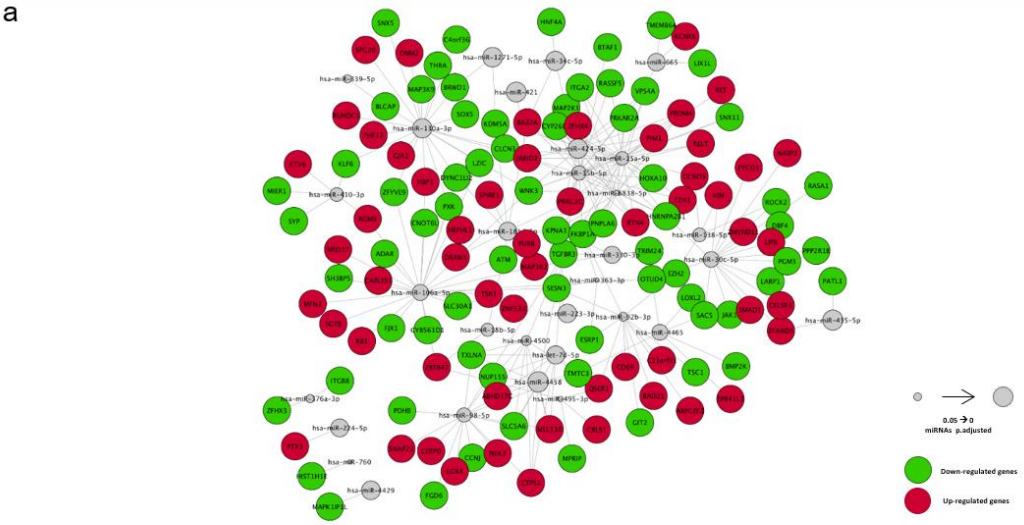
Figure 1. Pathway analysis during keloid formation. Hierarchical clustering of differentially activated pathways identified by the GSVA method for the differentially expressed genes specific to the healthy (top panel) and keloid-prone individuals (bottom panel). Green represents downregulated pathways and red represents overexpressed pathways.

Figure 2. miRNA-mRNA targetome and gene association network analysis during keloid formation. (a) Targetome of DE miRNAs. Size of grey nodes indicates the p-adjusted value for the differentially expressed miRNAs; the larger the node the more biological significant the miRNA is likely to have. Red nodes indicate upregulated genes while green nodes indicate downregulated genes in keloid subjects during keloid formation. **(b)** Gene association network during keloid formation (top panel). Genes are represented by colored nodes and are split into 27 clusters, sized by the number of genes in them. Genes in the same cluster are connected with black edges and genes that are connected across clusters with red edges. Average expression change in each cluster of the gene association network in keloid-prone and healthy individuals between baseline and 6 weeks after wounding (bottom panel). Each cluster is referenced by the x, y coordinate system. The x-axis represents time points, while the y-axis represents the average expression in each cluster. All expression levels are standardized and centered for plotting purposes.

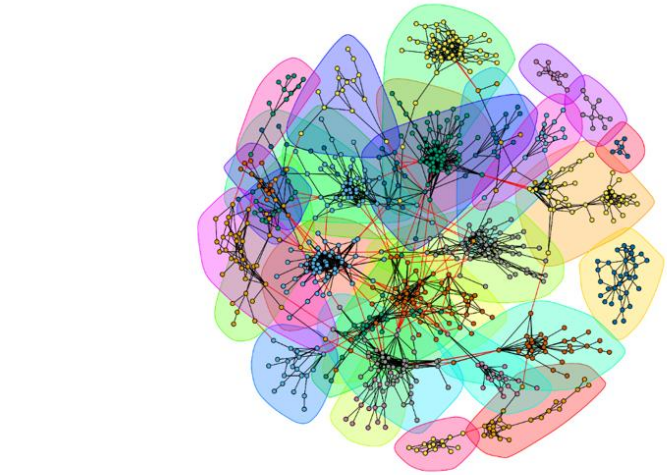

Control - Day 0
Control - Day 42
KIM WT1 TARGETS DN
HYPERTROPHIC CARDIOMYOPATHY HC
GINESTIER BREAST CANCER ZNF217
KIM WT1 TARGETS UP
CHEN HOXA5 TARGETS 9HR UP
NUYTTEN NIPP1 TARGETS UP
FULCHER INFLAMMATORY RESPONSE
LEE NEURAL CREST STEM CELL DN
CAIRO HEPATOBLASTOMA UP
JISON SICKLE CELL DISEASE UP
DAVICIONI TARGETS OF PAX FOXO1
BROWNE HCMV INFECTION 24HR DN
RODRIGUES THYROID CARCINOMA PO
RODRIGUES THYROID CARCINOMA AN
BROWNE HCMV INFECTION 4HR DN
TAKEDA TARGETS OF NUP98 HOXA9
TONKS TARGETS OF RUNX1 RUNX1T1
ENDOCYTOSIS
METABOLISM OF LIPIDS AND LIPOP
SPIELMAN LYMPHOBLAST EUROPEAN
NIKOLSKY BREAST CANCER 19P13 A
ELVIDGE HYPOXIA DN
DEBIASI APOPTOSIS BY REOVIRUS
HOSHIDA LIVER CANCER SURVIVAL
Gene sets and pathways
samples

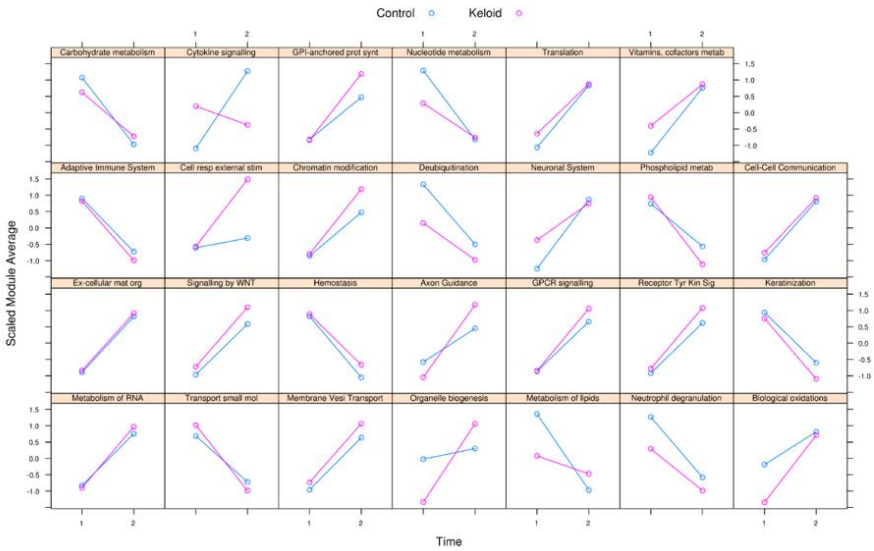

Keloid - Day 0
Keloid - Day 42
Gene sets and pathways
samples

a


0.05 → 0
miRNAs p.adjusted
Down-regulated genes
Up-regulated genes

b

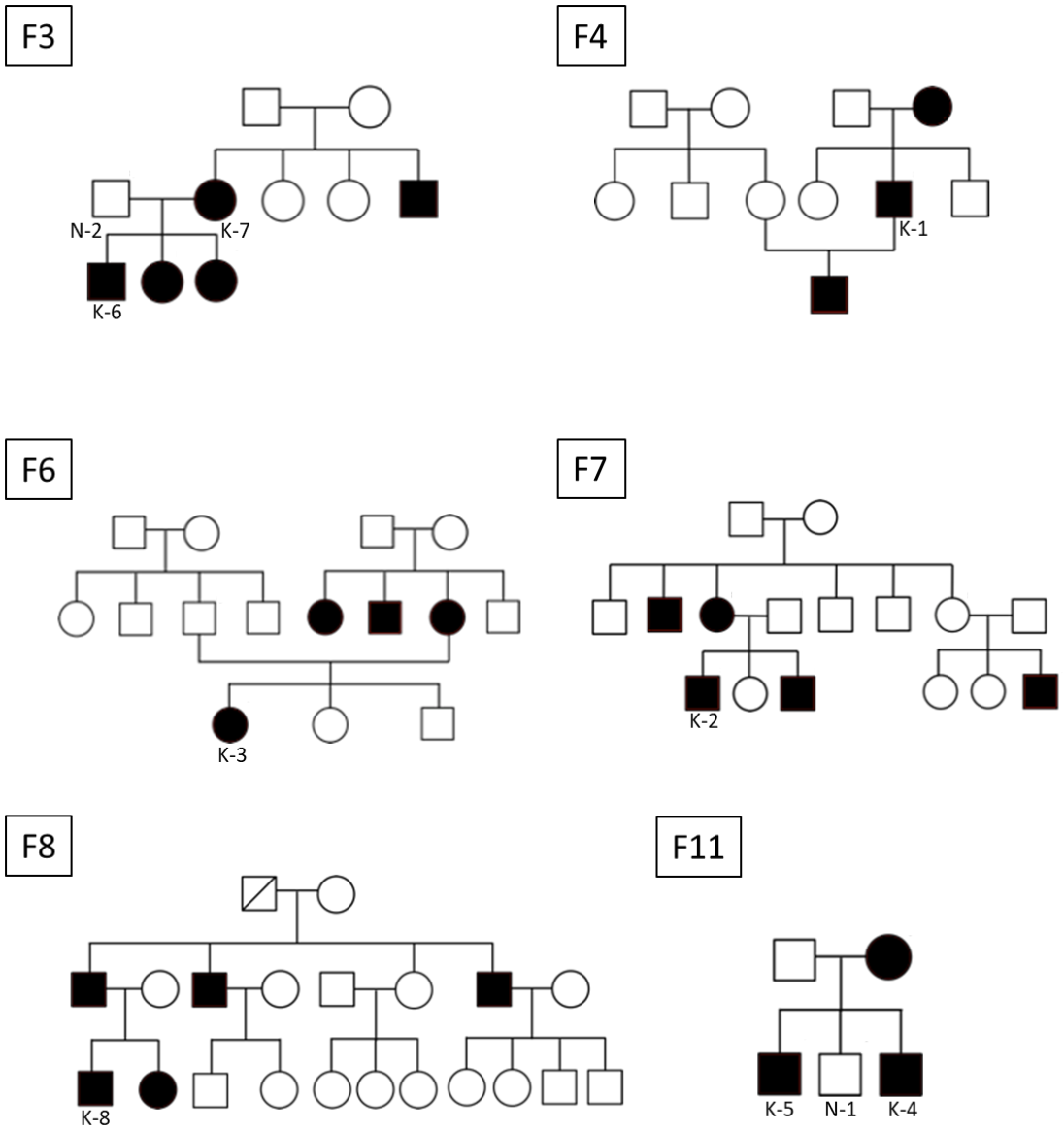



Control
Keloid
Carbohydrate metabolism
Cytokine signalling
GPI-anchored prot synt
Nucleotide metabolism
Translation
Vitamins, cofactors metab
Adaptive Immune System
Cell resp external stim
Chromatin modification
Deubiquitination
Neuronal System
Phospholipid metab
Cell-Cell Communication
Ex-cellular mat org
Signalling by WNT
Hemostasis
Axon Guidance
GPCR signalling
Receptor Tyr Kin Sig
Keratinization
Metabolism of RNA
Transport small mol
Membrane Vesi Transport
Organelle biogenesis
Metabolism of lipids
Neutrophil degranulation
Biological oxidations
Scaled Module Average
Time

Supplementary material

Time series integrative analysis of RNA-Seq and miRNA expression data reveals key biologic wound healing pathways in keloid-prone individuals

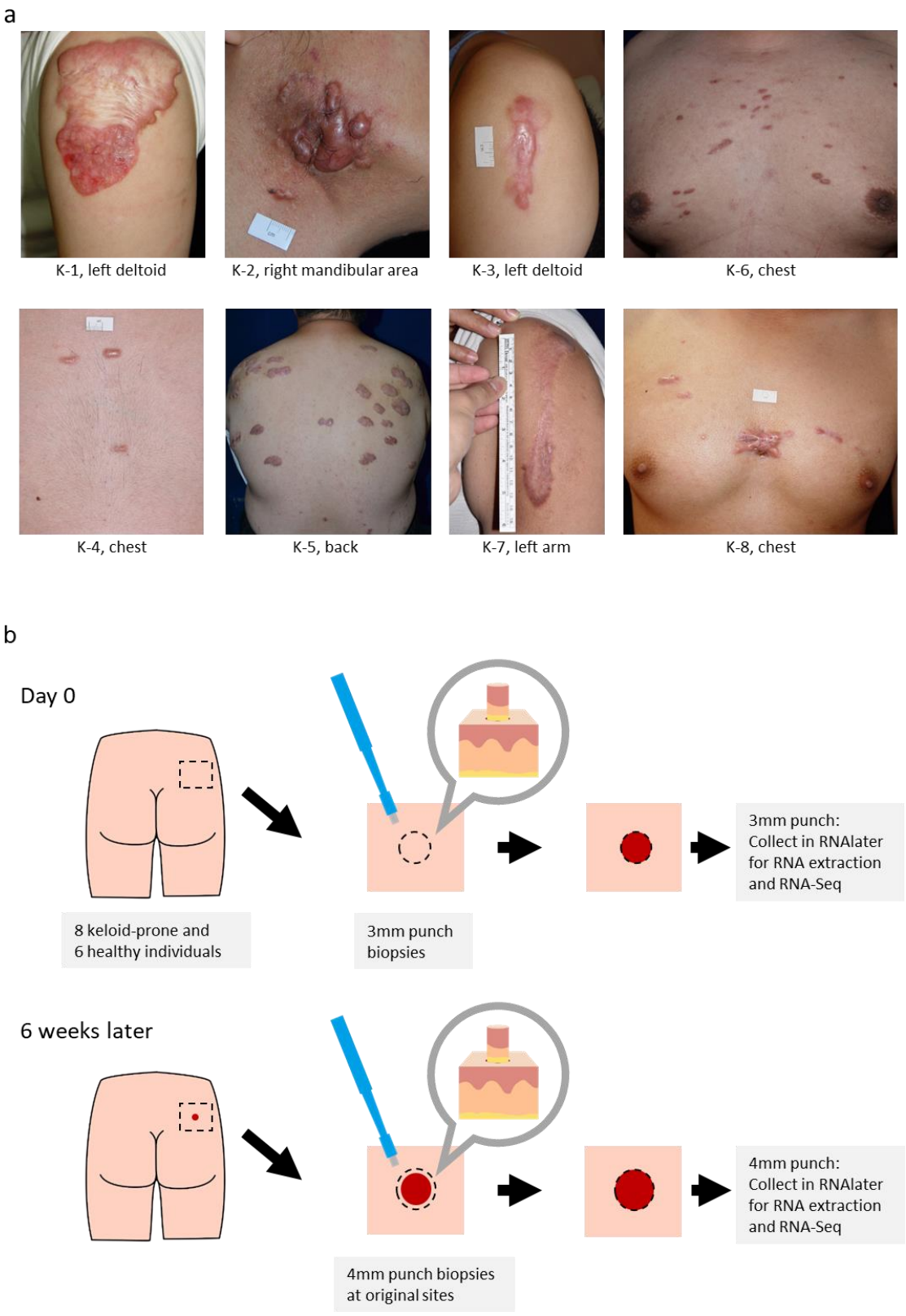
Alexandros Onoufriadis et al.



Supplementary Figure S1. Pedigrees from whom keloid-prone individuals were recruited for this study. Study skin biopsies were obtained from keloid-prone individuals indicated with a K and controls with N. An additional 4 matched controls not contained in these pedigrees were also included in the study. F numbers depict internal pedigree reference numbers.

Supplementary Table S1. Demographics for keloid-prone and control samples

Sample ID	Age (years)	Sex	Main keloid-prone areas	Duration (years)
Keloid				
K-1	32	male	bilateral arms, abdomen	10
K-2	23	male	bilateral mandibular area, chest	7
K-3	28	female	left arm	12
K-4	23	male	chest	5
K-5	24	male	back	12
K-6	30	male	chest, back	18
K-7	57	female	Left arm	44
K-8	25	male	chest	10
Control				
N-1	24	male		
N-2	58	male		
N-3	28	male		
N-4	22	male		
N-5	31	male		
N-6	34	male		



Supplementary Figure S2. Study design and keloid phenotype. (a) Illustration of established keloid lesions in the keloid-prone individuals who participated in the study. **(b)** Skin biopsy protocol and sample preparation for non-lesional skin in these individuals and the control subjects. The upper outer buttock was selected as the biopsy site in all individuals (keloid-prone and controls). The second punch biopsy was taken

directly overlying the initial punch biopsy. Following the second biopsy, the biopsy site in the keloid-prone individuals was injected with intra-dermal triamcinolone (10mg/ml) to try to reduce the risk of subsequent keloid scarring at this site. Nevertheless, within 3 months, 2 of 8 individuals went on to develop keloids at the biopsy site necessitating further intralesional triamcinolone treatment to reduce keloid size. None of the control subjects wound sites went on to form keloids.

Limitations of the study model.

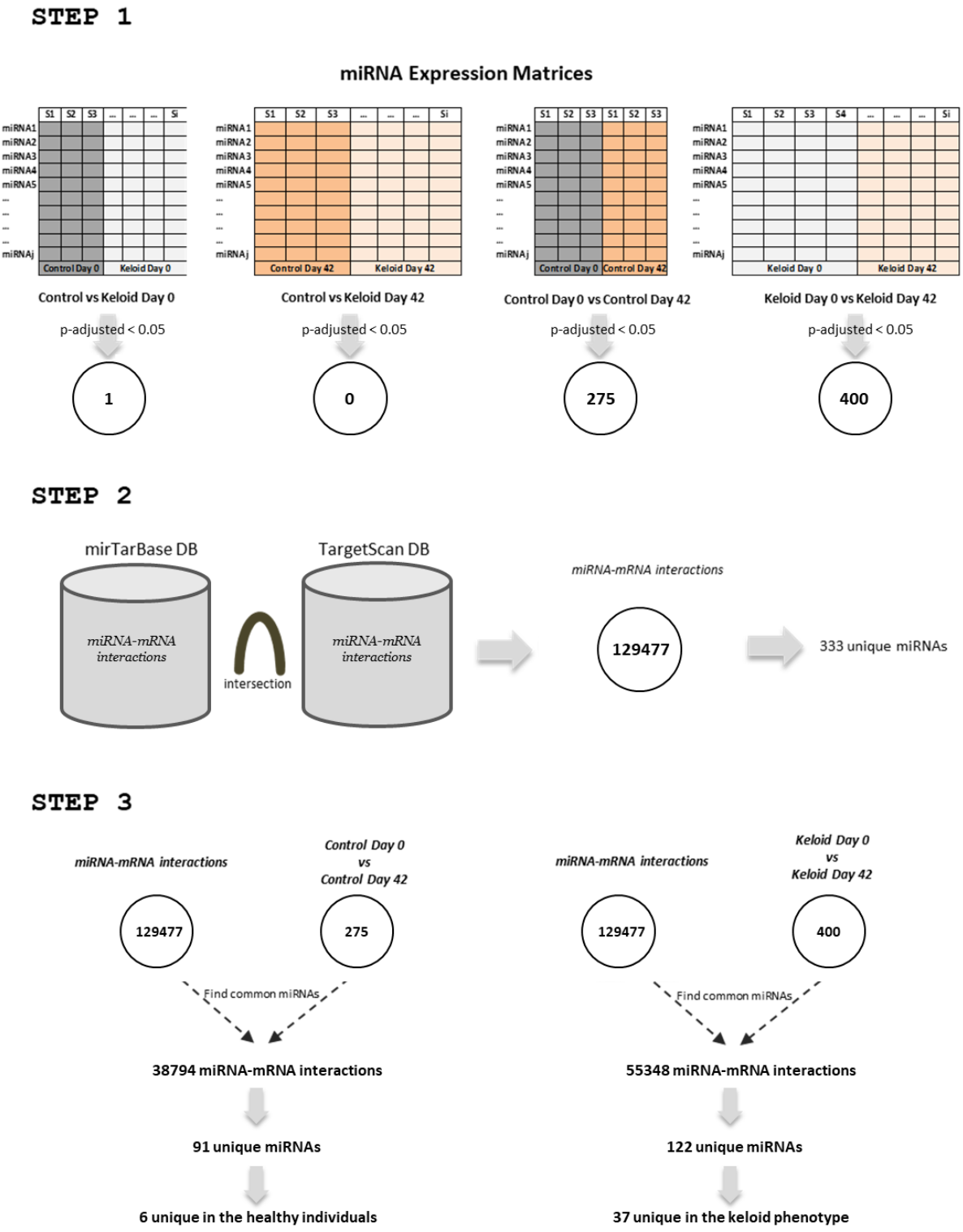
The buttock is not a prototypic site for keloid formation. However, it was not deemed ethically acceptable to biopsy more typical keloid-prone sites such as the skin overlying the sternum or shoulders. Nor were such biopsy sites acceptable to the keloid-prone participants (or control subjects). Ethics' committee approval also required us to try to avoid keloids forming after our second biopsies by injecting intra-lesional steroids to the biopsy sites, similar to what might be done in clinical practice in keloid-prone individual undergoing minor skin surgical procedures. The question therefore arises whether biopsying buttock skin is able to truly reflect the cellular pathology and events that typify what happens during "normal" keloid formation? The observation that two individuals in the study developed keloids at the biopsy site, despite the intra-lesional steroid injections, indicates that keloids - in keloid-prone individuals – can develop at this site, and that biopsying buttock skin may provide a useful and acceptable model for analyzing keloid biogenesis. Indeed, some of the other 6 keloid-prone individuals may also have gone on to develop keloids at the biopsy sites were it not for the injection of intra-lesional steroids, and thus undertaking subgroup analysis of those individuals who did or did not form keloids is not appropriate for this study. Nevertheless, it is important to reflect that our model for study may not precisely reflect events at other more typical keloid sites.

miRNA Expression Analysis

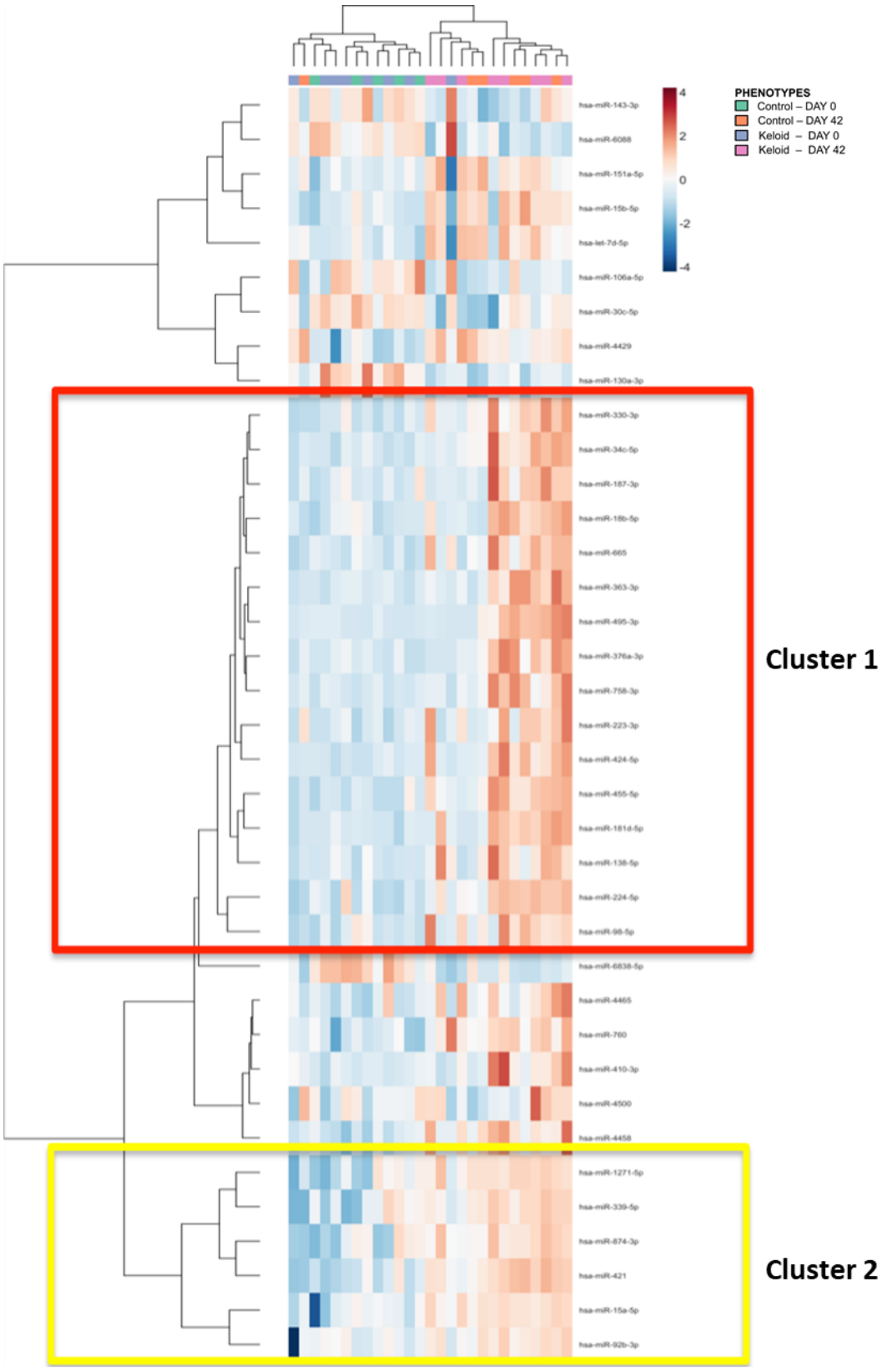
miRNA expression values were quantile normalized using Robust Multi-array Average (RMA) package in R and differential expression analysis was applied in the dataset across four different groups of samples (between keloid-prone and non-prone subjects at baseline and 6 weeks after wounding). Principal component analysis (PCA) was used to assess the clustering of samples. The limma (Linear Models for Microarray and RNA-Seq Data) package embedded in R (http://www.r-project.org/) was used to select miRNAs whose mean expression level was significantly different between different experimental conditions and further generated a list of microRNAs with associated statistics (Ritchie et al., 2015). Population level control and one factor analysis were used and miRNAs with an absolute fold change greater than 1.5 and adjusted *p*-value < 0.05 were selected as candidates that have significantly different expression for each comparison.

Initially, bioinformatics analysis identified 275 differentially expressed miRNAs in healthy individuals between baseline and 6 weeks after wounding, whereas 400 differentially expressed miRNAs were identified in the keloid-prone individuals between the two time points (Supplementary Figure S3 – Step 1). Next, the intersection of mirTarBase and TargetScanDB databases was used to annotate the differential expressed miRNAs across the four different conditions with their predicted target genes, resulting in 129,477 common miRNA-mRNA interactions in the two databases and 333 unique miRNAs (Supplementary Figure S3 – Step 2). By cross-referencing the differentially expressed miRNAs with the unique miRNAs of the miRNA-mRNA interactions, we identified 91 miRNAs that differentiate the healthy individuals between the two time points and 122 miRNAs for the keloid-prone individuals. Of

those, there were 6 miRNAs specific to the healthy individuals and 37 miRNAs specific to the keloid-prone subjects (Supplementary Figure S3 – Step 3).



Supplementary Figure S3. Stepwise differential expression profiling analysis of the miRNA expression data.



Supplementary Figure S4. Hierarchical clustering of the 37 unique differentially expressed miRNAs in the keloid phenotype, following cross-reference with the unique miRNAs of the miRNA-mRNA interaction databases. Two clusters discriminate miRNAs that exhibit an upregulation in keloid prone subjects 6 weeks after wounding.

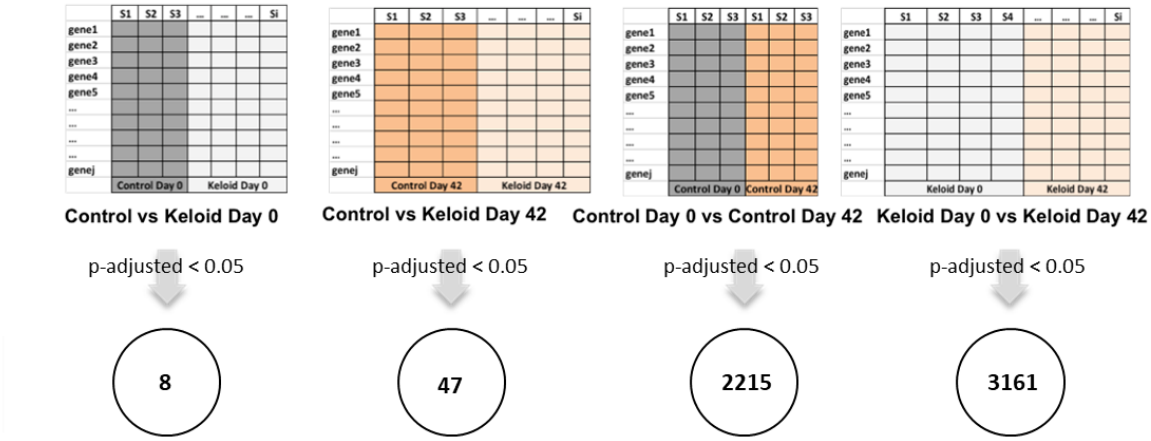
RNA-Seq library preparation and sequencing

Whole transcriptome RNA-Seq libraries were prepared on the Agilent Bravo liquid handling system using a modification of the Agilent Sureselect stranded RNA kit (Catalog 9691B). Briefly, ribosomal RNA was removed from 900ng of high quality RNA using the Ribozero Gold rRNA removal kit (Illumina, Cambridge cat MRZG12324) and ribodepletion was verified on a Bioanalyzer Pico RNA chip (Agilent catalog 5017-1513). Ribodepleted RNA was fragmented and reverse transcribed with random primers in the presence of Actinomycin D (Sigma) to inhibit antisense artifacts through DNA dependent DNA synthesis (Perocchi et al., 2007). Second strand was synthesized with dUTP, the ends of ds-cDNA were polished followed by adaptor ligation and PCR amplification in the presence of Uracil-DNA-Glycosylase (UDG) to selectively degrade the second strand. Libraries were bead purified using Ampure beads, fragment size confirmed on Agilent Tapestation D1000 Screentape and quantified with Qubit. Sequencing was conducted on HiSeq2500 and HiSeq3000 (Illumina, Cambridge).

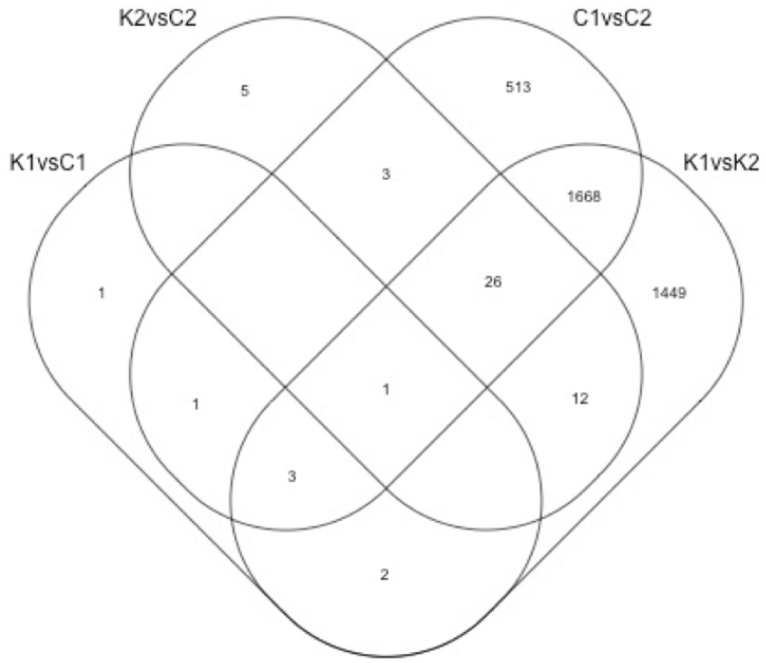
RNA-Seq Data Analysis

Sequencing reads were aligned against the human genome sequence (hg38) using hisat2. Reads with a MAPQ (mapping quality) below 10 and PCR duplicate reads were removed with Samtools package. On average, 121 million reads per samples were obtained. The UCSC hg38 genome annotation was used to generate gene count tables for each sample using the GenomicAlignments library in R (Lawrence et al., 2013). Gene counts for the technical replicates were added together and data were normalized for library size correction using DESeq2. Principal Component Analysis (PCA) on the normalized gene counts was performed to ensure the homogeneity of the data and to exclude any potential batch effects among sequencing runs. Sample N3-2nd was removed because of outlier status. Following normalization, low-expressed genes with an average expression across all samples of less than 3 reads were removed. RNA differential expression (DE) analysis was performed with the glmmADMB package in R using a mixed effect generalized linear model with a random intercept for each patient from the negative binomial family (Fournier et al. 2012).

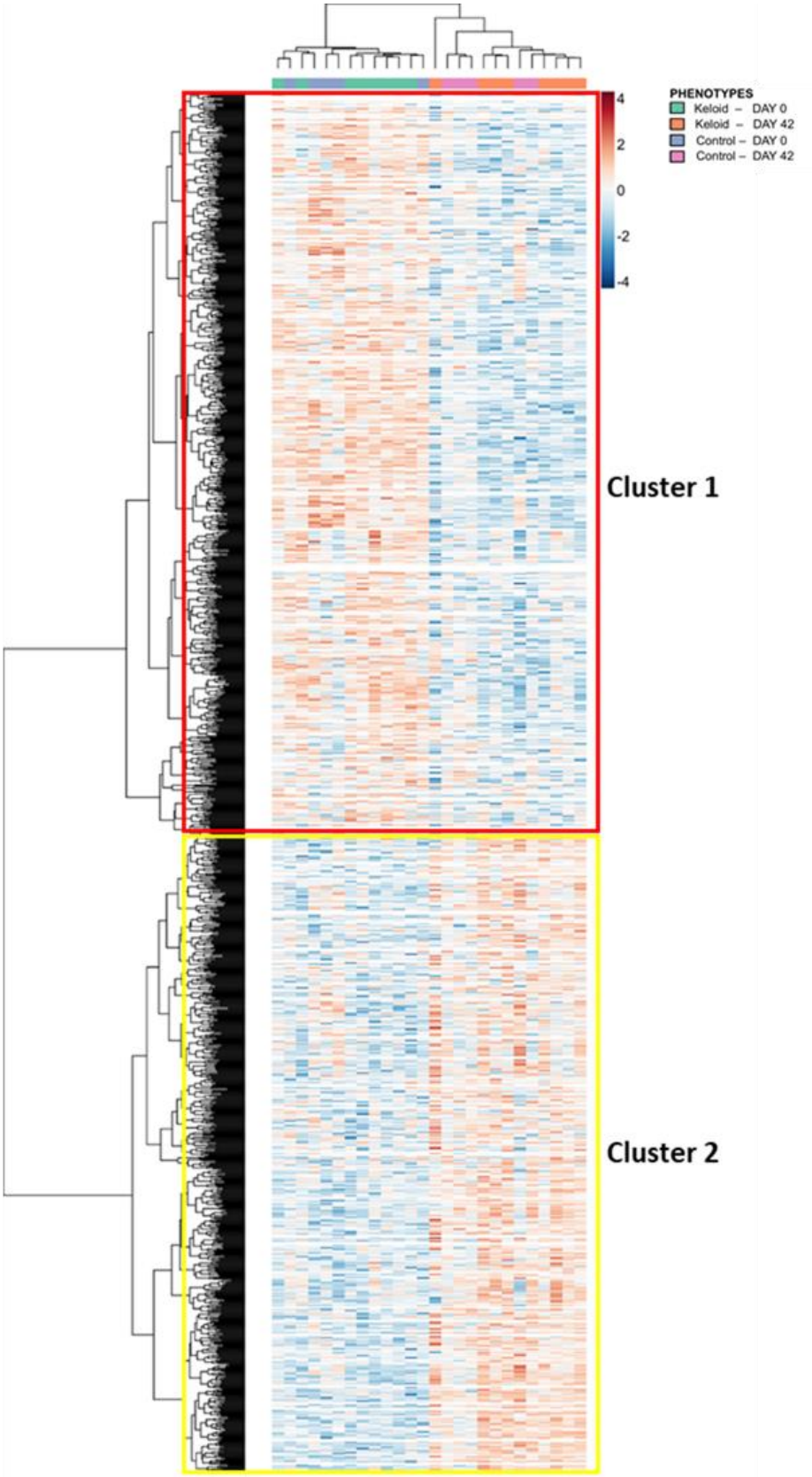
a



b



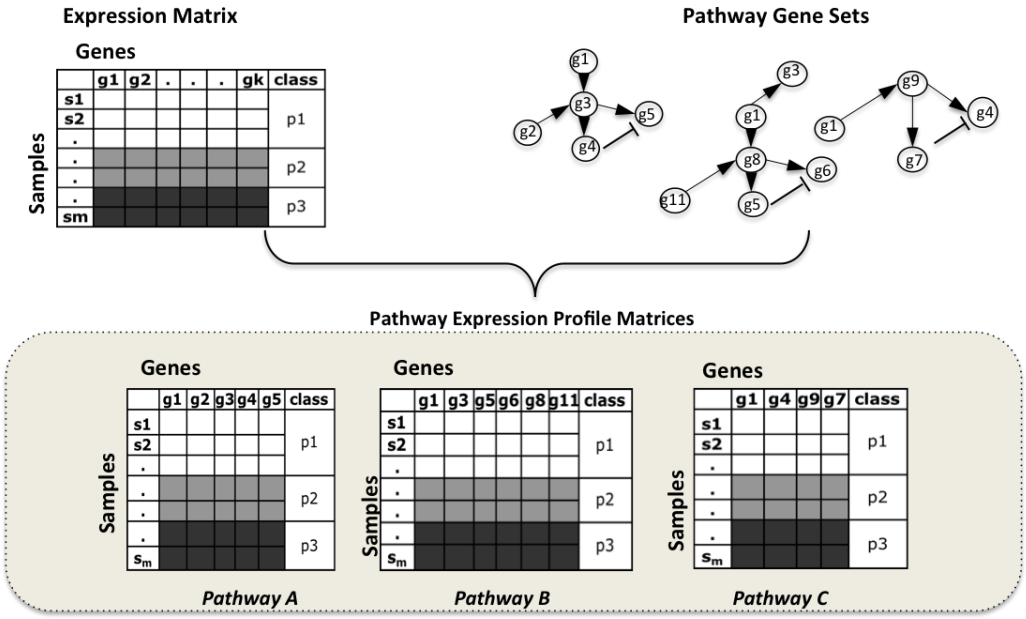
Supplementary Figure S5. (a) Differentially expressed genes across the different phenotypes before and after wounding. **(b)** Venn diagram reveals unique genes in the pairwise comparisons.



Supplementary Figure S6. Hierarchical clustering reveals two distinct clusters with expression differences between baseline and 6 weeks after wounding for the 1449 keloid-specific genes.

Pathway Enrichment Analysis

Gene Set Variation Analysis (GSVA) of the DE RNA-Seq data was performed in R using the C2 collection of curated gene sets that form part of the Molecular Signatures Database (MSigDB) version 3.0. GSVA transforms the data from a gene-by-sample matrix to a gene set-by-sample matrix by performing a change in coordinate systems, and therefore allowing for the evaluation of pathway enrichment for each sample (Supplementary Figure S7). A stringent cut-off (adjPvalueCutoff = 0.001) was employed to attain high level of statistical and biologic significance.



Supplementary Figure S7. Schematic representation of the GSVA method showing the conversion of a gene-by-sample matrix to a gene set-by-sample matrix.

Supplementary Table S2. Differentially regulated pathways identified by the GSVA method for the 513 genes that were specific to the healthy individuals.

Gene Sets (according to MSigDB C2 collection)	logFC*	AveExpr	t	P.Value	adj.P.Val	B
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_DN	1.183782	-0.02015	10.05267	4.44E-09	1.34E-06	11.09738
BROWNE_HCMV_INFECTION_4HR_DN	1.169318	-0.00657	9.712124	7.75E-09	1.34E-06	10.55167
NIKOLSKY_BREAST_CANCER_16P13_AMPLICON	-1.03877	-0.00717	-8.64669	4.84E-08	5.58E-06	8.747387
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_3D_UP	1.099515	0.011772	8.319773	8.73E-08	7.55E-06	8.163094
RODRIGUES_THYROID_CARCINOMA_ANAPLASTIC_UP	0.849148	-0.00568	7.090821	9.12E-07	5.83E-05	5.832659
RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_UP	0.882856	-0.00633	7.038979	1.01E-06	5.83E-05	5.729668
SPIELMAN_LYMPHOBLAST_EUROPEAN_VS_ASIAN_UP	-0.83025	-0.00941	-6.92105	1.28E-06	6.33E-05	5.493992
ELVIDGE_HYPOXIA_DN	-0.98957	0.02289	-6.44432	3.40E-06	0.000131	4.521731
CAIRO_HEPATOBLASTOMA_UP	0.807335	-0.02311	6.441304	3.42E-06	0.000131	4.515492
GINESTIER_BREAST_CANCER_ZNF217_AMPLIFIED_DN	0.83585	0.019098	6.045407	7.86E-06	0.000272	3.684998
BROWNE_HCMV_INFECTION_24HR_DN	0.887437	0.02928	5.770285	1.42E-05	0.000432	3.096397
JISON_SICKLE_CELL_DISEASE_UP	0.785883	0.036665	5.725194	1.56E-05	0.000432	2.999086
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	0.739191	-0.04794	5.685418	1.71E-05	0.000432	2.913056
REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS	-0.80359	-0.01124	-5.67427	1.75E-05	0.000432	2.888906
CHEN_HOXA5_TARGETS_9HR_UP	0.799507	0.043153	5.596677	2.07E-05	0.000477	2.720493
KEGG_ENDOCYTOSIS	-0.85191	-0.01476	-5.35382	3.53E-05	0.000763	2.189317
FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS_DN	0.667295	0.014279	5.278539	4.17E-05	0.000828	2.0235
LEE_NEURAL_CREST_STEM_CELL_DN	0.723797	0.037029	5.263428	4.31E-05	0.000828	1.990152
KIM_WT1_TARGETS_UP	0.664411	0.002188	5.193463	5.03E-05	0.000904	1.835495
HOSHIDA_LIVER_CANCER_SURVIVAL_DN	-0.70418	-0.01376	-5.14383	5.62E-05	0.000904	1.725532
DEBIASI_APOPTOSIS_BY_REOVIRUS_INFECTION_DN	-0.72564	0.014966	-5.13046	5.79E-05	0.000904	1.69588
KIM_WT1_TARGETS_DN	0.607918	-0.02711	5.121005	5.92E-05	0.000904	1.67489
NUYTTEN_NIPP1_TARGETS_UP	0.663922	0.014935	5.103511	6.15E-05	0.000904	1.636052
DAVICIONI_TARGETS_OF_PAX_FOXO1_FUSIONS_UP	0.731321	0.053744	5.095033	6.27E-05	0.000904	1.617221

* FC refers to the differential expression between control samples at day 42 vs Day 0
For MSigDB collections:
http://software.broadinstitute.org/gsea/msigdb/collections.jsp

Supplementary Table S3. Differentially regulated pathways identified by the GSVA method for the 1449 genes that were specific to the keloid-prone individuals.

Gene Sets* (according to MSigDB C2 collection)	logFC**	AveExpr	t	P.Value	adj.P.Val	B
KOBAYASHI_EGFR_SIGNALING_24HR_UP	1.000138	-0.01189	10.55765	3.62E-10	1.44E-07	13.44671
LAU_APOPTOSIS_CDKN2A_UP	0.993493	-0.01853	10.86145	2.12E-10	1.44E-07	13.96596
REACTOME_DNA_REPAIR	-0.94955	-0.00332	-10.177	7.20E-10	1.79E-07	12.77967
TURASHVILI_BREAST_DUCTAL_CARCINOMA_VS_LOBULAR_NORMAL_UP	1.058949	-0.01511	10.05451	9.01E-10	1.79E-07	12.561
CHEMNITZ_RESPONSE_TO_PROSTAGLANDIN_E2_UP	-1.0039	0.014287	-9.87279	1.26E-09	2.01E-07	12.23302
KEGG_NOTCH_SIGNALING_PATHWAY	-0.98173	0.043333	-9.71578	1.69E-09	2.14E-07	11.94611
SABATES_COLORECTAL_ADENOMA_DN	0.850325	-0.00294	9.658409	1.89E-09	2.14E-07	11.84046
NIKOLSKY_BREAST_CANCER_16P13_AMPLICON	-0.97521	0.013771	-9.24433	4.17E-09	4.14E-07	11.0647
KEGG_MAPK_SIGNALING_PATHWAY	-0.82233	-0.02534	-8.96403	7.22E-09	4.73E-07	10.52623
MAYBURD_RESPONSE_TO_L663536_DN	-0.90647	0.013188	-9.02439	6.41E-09	4.73E-07	10.64311
REACTOME_INNATE_IMMUNITY_SIGNALING	-0.9752	0.012237	-8.92926	7.74E-09	4.73E-07	10.45869
REACTOME_TOLL_LIKE_RECEPTOR_3_CASCADE	-0.9752	0.012237	-8.92926	7.74E-09	4.73E-07	10.45869
REACTOME_TOLL_RECEPTOR_CASCADES	-0.9752	0.012237	-8.92926	7.74E-09	4.73E-07	10.45869
BROWNE_HCMV_INFECTION_6HR_UP	-0.98553	-0.019	-8.75839	1.09E-08	6.18E-07	10.12426
SCHUHMACHER_MYC_TARGETS_UP	-0.97424	0.001571	-8.59373	1.52E-08	8.03E-07	9.798136
ODONNELL_TFRC_TARGETS_DN	-0.80427	-0.02557	-8.5514	1.65E-08	8.20E-07	9.713684
REACTOME_GLUCOSE_REGULATION_OF_INSULIN_SECRETION	-0.90846	-0.00318	-8.25928	3.01E-08	1.41E-06	9.123977
MISSIAGLIA_REGULATED_BY_METHYLATION_DN	-0.90167	0.010577	-8.10898	4.11E-08	1.81E-06	8.815857
KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM	-0.90756	-0.00807	-8.05696	4.58E-08	1.87E-06	8.70848
RODRIGUES_DCC_TARGETS_DN	0.716196	0.005976	8.043151	4.71E-08	1.87E-06	8.679911
CHENG_IMPRINTED_BY_ESTRADIOL	1.043827	0.008587	7.946755	5.78E-08	2.01E-06	8.479728
OSWALD_HEMATOPOIETIC_STEM_CELL_IN_COLLAGEN_GEL_DN	-0.76098	-0.01989	-7.94314	5.82E-08	2.01E-06	8.472189
OSWALD_HEMATOPOIETIC_STEM_CELL_IN_COLLAGEN_GEL_UP	-0.76098	-0.01989	-7.94314	5.82E-08	2.01E-06	8.472189
PODAR_RESPONSE_TO_ADAPHOSTIN_UP	0.729459	0.013285	7.854346	7.02E-08	2.33E-06	8.286596
CHIANG_LIVER_CANCER_SUBCLASS_INTERFERON_DN	-0.77226	-0.04696	-7.81695	7.60E-08	2.42E-06	8.208096
WEST_ADRENOCORTICAL_TUMOR_DN	0.563721	-0.02509	7.765676	8.48E-08	2.59E-06	8.100147
REACTOME_INTEGRATION_OF_ENERGY_METABOLISM	-0.81238	-0.01477	-7.70488	9.66E-08	2.84E-06	7.971668
CHIANG_LIVER_CANCER_SUBCLASS_CTNNB1_DN	0.848678	0.037433	7.430023	1.75E-07	4.97E-06	7.38439
AMUNDSON_POOR_SURVIVAL_AFTER_GAMMA_RADIATION_8G	-0.88049	-0.00325	-7.35071	2.08E-07	5.68E-06	7.212967
CHAUHAN_RESPONSE_TO_METHOXYESTRADIOL_DN	0.807922	0.004595	7.33706	2.14E-07	5.68E-06	7.183385
BENPORATH_ES_1	-0.59604	0.00292	-7.20347	2.88E-07	7.31E-06	6.892458
NGUYEN_NOTCH1_TARGETS_DN	-0.90911	0.010124	-7.19338	2.94E-07	7.31E-06	6.870381
KAUFFMANN_DNA_REPLICATION_GENES	-0.74468	0.029854	-7.05786	3.98E-07	8.95E-06	6.572588
PUJANA_BREAST_CANCER_LIT_INT_NETWORK	-0.71923	0.027218	-7.06725	3.89E-07	8.95E-06	6.5933
REACTOME_REGULATION_OF_INSULIN_SECRETION	-0.8035	-0.00883	-7.083	3.76E-07	8.95E-06	6.628031
WEI_MIR34A_TARGETS	-0.73811	0.010186	-7.04957	4.05E-07	8.95E-06	6.554288
FIRESTEIN_PROLIFERATION	-0.70045	0.025921	-7.01333	4.39E-07	9.19E-06	6.474208
REACTOME_AXON_GUIDANCE	-0.65862	-0.02215	-7.01493	4.38E-07	9.19E-06	6.477755
VECCHI_GASTRIC_CANCER_EARLY_UP	-0.61891	0.004229	-6.95025	5.06E-07	1.03E-05	6.334388
LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_DN	0.731221	-0.03316	6.890709	5.78E-07	1.15E-05	6.201912

ASTON_MAJOR_DEPRESSIVE_DISORDER_DN	-0.63798	0.009268	-6.86316	6.15E-07	1.19E-05	6.140469
RUIZ_TNC_TARGETS_DN	-0.72356	0.005279	-6.81771	6.82E-07	1.29E-05	6.038882
LINDGREN_BLADDER_CANCER_CLUSTER_1_DN	-0.60063	-0.00917	-6.70501	8.80E-07	1.63E-05	5.785799
SONG_TARGETS_OF_IE86_CMV_PROTEIN	-0.82487	0.006249	-6.63537	1.03E-06	1.86E-05	5.628616
PUIFFE_INVASION_INHIBITED_BY_ASCITES_DN	-0.67083	0.008931	-6.57148	1.19E-06	2.11E-05	5.483868
VERHAAK_AML_WITH_NPM1_MUTATED_UP	-0.60631	-0.02712	-6.51353	1.36E-06	2.36E-05	5.352154
PELLICCIOTTA_HDAC_IN_ANTIGEN_PRESENTATION_UP	-0.68774	-0.00684	-6.42971	1.65E-06	2.80E-05	5.160903
ROSS_AML_WITH_MLL_FUSIONS	-0.85834	-0.01609	-6.31243	2.17E-06	3.60E-05	4.891892
MASSARWEH_TAMOXIFEN_RESISTANCE_DN	-0.59446	-0.02214	-6.25596	2.48E-06	4.02E-05	4.761778
BENPORATH_PRC2_TARGETS	-0.59854	-0.00483	-6.18979	2.89E-06	4.59E-05	4.608869
REACTOME_BIOLOGICAL_OXIDATIONS	0.83103	0.009525	6.076609	3.77E-06	5.76E-05	4.346168
REACTOME_PHASE_1_FUNCTIONALIZATION_OF_COMPOUNDS	0.83103	0.009525	6.076609	3.77E-06	5.76E-05	4.346168
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	0.510054	0.008107	6.033497	4.17E-06	6.25E-05	4.24574
PROVENZANI_METASTASIS_DN	-0.62241	-0.01288	-5.89672	5.76E-06	8.33E-05	3.92584
WILCOX_PRESPONSE_TO_ROGESTERONE_UP	0.56312	0.022466	5.896236	5.77E-06	8.33E-05	3.924704
KIM_WT1_TARGETS_DN	0.487331	-0.00753	5.836371	6.65E-06	9.27E-05	3.78409
WONG_EMBRYONIC_STEM_CELL_CORE	-0.57525	-0.00298	-5.83926	6.60E-06	9.27E-05	3.790888
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	-0.59493	0.005355	-5.81028	7.07E-06	9.69E-05	3.722708
NIKOLSKY_BREAST_CANCER_20Q12_Q13_AMPLICON	0.542719	0.041836	5.753972	8.09E-06	0.000109	3.589984
RUTELLA_RESPONSE_TO_CSF2RB_AND_IL4_UP	0.529805	0.002341	5.694893	9.31E-06	0.000121	3.450425
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_6HR_UP	-0.66729	0.001285	-5.69535	9.30E-06	0.000121	3.451505
ZHONG_RESPONSE_TO_AZACITIDINE_AND_TSA_DN	-0.54356	0.013731	-5.68567	9.52E-06	0.000122	3.428609
REACTOME_PLATELET_ACTIVATION	0.531231	-0.00775	5.656474	1.02E-05	0.000128	3.359502
WINNEPENNINCKX_MELANOMA_METASTASIS_UP	-0.59367	0.003779	-5.6538	1.03E-05	0.000128	3.353162
REACTOME_METABOLISM_OF_PROTEINS	0.619841	0.033359	5.637109	1.07E-05	0.000131	3.313624
WONG_ENDMETRIUM_CANCER_DN	0.75604	-0.01179	5.594264	1.18E-05	0.000143	3.212005
KIM_WT1_TARGETS_12HR_UP	-0.73606	-0.0411	-5.58202	1.22E-05	0.000145	3.182928
KEGG_P53_SIGNALING_PATHWAY	-0.57218	-0.01645	-5.52567	1.40E-05	0.000163	3.048993
CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_5	-0.53053	-0.00105	-5.44002	1.72E-05	0.000196	2.844939
NIKOLSKY_BREAST_CANCER_17Q21_Q25_AMPLICON	0.423658	0.000168	5.437157	1.73E-05	0.000196	2.838105
FURUKAWA_DUSP6_TARGETS_PCI35_UP	0.54367	0.030323	5.40184	1.88E-05	0.000211	2.753797
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	-0.605	0.001353	-5.32512	2.27E-05	0.00025	2.570352
BILBAN_B_CLL_LPL_UP	0.630854	-0.016	5.307394	2.37E-05	0.000258	2.527907
SABATES_COLORECTAL_ADENOMA_UP	-0.6578	-0.00434	-5.23944	2.79E-05	0.0003	2.365021
DAVICIONI_TARGETS_OF_PAX_FOXO1_FUSIONS_DN	-0.67133	0.012634	-5.22906	2.86E-05	0.000303	2.340114
DANG_MYC_TARGETS_UP	-0.51829	-0.01363	-5.2015	3.06E-05	0.00032	2.273965
REACTOME_CELL_JUNCTION_ORGANIZATION	0.513445	-0.00263	5.183873	3.20E-05	0.00033	2.231629
WANG_SMARCE1_TARGETS_UP	0.627323	-0.02644	5.168352	3.32E-05	0.000338	2.19434
MUELLER_PLURINET	-0.55763	0.001991	-5.13036	3.64E-05	0.000366	2.103006
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	-0.61869	0.000997	-5.04602	4.47E-05	0.000444	1.900008
REACTOME_METABOLISM_OF_RNA	-0.60005	0.005003	-5.01078	4.87E-05	0.000478	1.815081
BENPORATH_PROLIFERATION	-0.4765	-0.01437	-4.9699	5.39E-05	0.000522	1.716523
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_16D_UP	-0.55114	0.020418	-4.94582	5.71E-05	0.000547	1.65845
BENPORATH_EED_TARGETS	-0.44183	-0.01099	-4.9147	6.17E-05	0.000584	1.583332
REACTOME_MUSCLE_CONTRACTION	0.692889	-0.03459	4.908079	6.27E-05	0.000586	1.567363

LEE_DIFFERENTIATING_T_LYMPHOCYTE	0.633821	-0.01295	4.807104	8.03E-05	0.000742	1.323492
SASSON_RESPONSE_TO_FORSKOLIN_DN	0.717346	0.029462	4.792682	8.32E-05	0.000752	1.288643
SASSON_RESPONSE_TO_GONADOTROPHINS_DN	0.717346	0.029462	4.792682	8.32E-05	0.000752	1.288643
REACTOME_FORMATION_OF_PLATELET_PLUG	0.415674	-0.00951	4.784697	8.48E-05	0.000758	1.269346
GAL_LEUKEMIC_STEM_CELL_UP	0.497293	0.010221	4.780143	8.58E-05	0.000758	1.258339
ELVIDGE_HYPOXIA_BY_DMOG_DN	0.617879	0.030054	4.766073	8.88E-05	0.000776	1.224333
CASORELLI_ACUTE_PROMYELOCYTIC_LEUKEMIA_DN	-0.40155	0.001137	-4.74356	9.39E-05	0.000811	1.169915
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_MONOCYTE_UP	-0.60661	0.03329	-4.72037	9.94E-05	0.000849	1.113856
BROWNE_HCMV_INFECTION_20HR_DN	0.685218	-0.02484	4.696294	0.000105	0.000892	1.055648
HAHTOLA_MYCOSIS_FUNGOIDES_CD4_DN	-0.62102	-0.00789	-4.67604	0.000111	0.000927	1.00668
BENPORATH_OCT4_TARGETS	-0.42012	0.003577	-4.65044	0.000118	0.000938	0.944783
RODWELL_AGING_KIDNEY_UP	0.458909	0.018979	4.661725	0.000115	0.000938	0.972064
SMID_BREAST_CANCER_ERBB2_UP	0.580798	0.020327	4.658416	0.000116	0.000938	0.964063
TOYOTA_TARGETS_OF_MIR34B_AND_MIR34C	-0.4227	-0.02818	-4.65284	0.000117	0.000938	0.950574
YEGNASUBRAMANIAN_PROSTATE_CANCER	-0.58234	-0.00018	-4.66109	0.000115	0.000938	0.970529
REN_ALVEOLAR_RHABDOMYOSARCOMA_DN	0.401401	-0.01494	4.64276	0.00012	0.000947	0.926206

* Gene sets present in the Reactome and KEGG databases are highlighted in yellow and orange respectively
** FC refers to the differential expression between keloid samples at day 42 vs Day 0
For MSigDB collections:
http://software.broadinstitute.org/gsea/msigdb/collections.jsp

Supplementary Table S4. Differentially activated pathways identified by the GSEA method for the 1,449 genes that were specific to the keloid-prone individuals.

	p.geomean	stat.mean	p.val	q.val
hsa04010 MAPK signaling pathway	0.10390259	-1.271314	0.00031	0.002787

Supplementary Table S5. MAPK signaling pathway genes involved in the keloid interactome between the RNA-Seq and miRNA datasets.

genes	miRNA	logfc.genes	pval.genes	p.adj.genes	p.adj.miRNAs	reg.genes	gs	ENSEMBL ID
MAP2K1	hsa-miR-34c-5p	-0.35476633	0.00024254	0.001315459	0.002911087	down	5604	ENSG00000169032
MAP2K1	hsa-miR-424-5p	-0.35476633	0.00024254	0.001315459	0.002641301	down	5604	ENSG00000169032
MAP3K2	hsa-miR-106a-5p	0.39859529	0.00682168	0.023968285	0.012701839	up	10746	ENSG00000169967
MAP3K2	hsa-miR-330-3p	0.39859529	0.00682168	0.023968285	0.009208344	up	10746	ENSG00000169967
RASA1	hsa-miR-30c-5p	-0.2833839	0.00019165	0.001060305	0.013428058	down	5921	ENSG00000145715

Gene Association Network Analysis

Differentially expressed genes from the four comparisons with adjusted P-value of less than 0.05 were used to construct gene association networks. Briefly, differentially expressed genes were represented in the form of a graph - where two genes are associated (i.e. were connected by a line in the graph) if they shared a Reactome ID (Croft et al., 2014) and their expression profiles were correlated above absolute value of 0.6. In this network of 757 associated genes, clusters were detected using edge betweenness community detection algorithm (Newman et al., 2004), and implemented in igraph (Csardi et al., 2006) library in R. Each cluster is assigned a class label which is the most common Reactome term shared among the member genes of the respective cluster. The average expression profile of the genes in each cluster is calculated and plotted.

Supplementary References

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014;42:D472-7.

Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 2006;1695.

Fournier DA, Skaug HJ, Ancheta J, Ianelli J, Magnusson A, Maunder MN, et al. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 2012;27:233–249.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9:e1003118.

Newman ME, Girvan M. Finding and evaluating community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2004;69:026113.

Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Res. 2007;35:e128.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.